

**Penerapan Data Mining dengan Menggunakan Metode Clustering
K-Mean Untuk Mengukur Tingkat Ketepatan Kelulusan
Mahasiswa Program Teknik Informatika S1
Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang
Tahun Ajaran 2012/2013**

RIZAL AKBAR

*Program Studi Teknik Informatika – S1, Fakultas Ilmu Komputer, Universitas
Dian Nuswantoro*

Jl. Imam Bonjol 207 Semarang 50131

Telp. (024) 3517261, Fax (024) 3520165

URL : <http://dinus.ac.id/>, email : rizalandroid@gmail.com

ABSTRAK

Faktor-faktor yang mempengaruhi mahasiswa mengalami kemunduran masa studi yaitu mahasiswa yang selama beberapa semester berturut-turut mendapatkan nilai atau IP semester rendah atau kurang dari standar yang ditetapkan oleh akademi, serta tidak dapat memenuhi ketentuan masa studi,

Dengan pengelompokkan mahasiswa secara manual berdasarkan nilai rata-rata mereka, maka tidak mudah untuk memperoleh pandangan yang menyeluruh mengenai keadaan nilai prestasi mahasiswa angkatan 2012. Adapun tujuan penelitian ini adalah untuk melakukan pengelompokan tentang jumlah mahasiswa mengenai waktu kuliah pada program studi Teknik Informatika Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang Tahun Ajaran 2012/2013.

Dengan pengelompokan masa studi mahasiswa Teknik Informatika 2012 maka pihak universitas dapat melakukan tindakan atau membuat kebijakan untuk mengatasi permasalahan kemunduran masa studi. Data mahasiswa yang berpotensi mengalami kemunduran 132 mahasiswa dan mahasiswa yang berpotensi menyelesaikan masa studinya lebih cepat 156 mahasiswa dan data Mahasiswa yang berpotensi menyelesaikan masa studinya sesuai dengan standar adalah 225 Mahasiswa

Dari data tersebut apabila digali dengan tepat maka dapat diketahui pola atau pengetahuan baru yang dapat dikembangkan untuk diterapkan pada perguruan tinggi diantaranya mengenai potensi mahasiswa *drop out*.

Pemahaman informasi mahasiswa yang potensial terkena *drop out* ini penting untuk diketahui dan dipahami.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Perguruan tinggi adalah salah satu institusi yang sudah pasti memiliki data yang tidak kecil volumenya. Database perguruan tinggi menyimpan berbagai data, misalnya data akademik, data administrasi dan data mahasiswa.

Pemahaman dapat dilakukan dengan menggali data – data yang dimiliki dan kemudian dilakukan pengelompokan terhadap hasil penggalian data tersebut sehingga memunculkan pola atau kelompok mahasiswa yang berpotensi terkena *drop out* .

Pencegahan kegagalan adalah sangat penting bagi manajemen perguruan tinggi. Pengetahuan ini dapat digunakan dalam membantu pihak perguruan tinggi untuk lebih mengenal situasi para mahasiswanya dan dapat dijadikan sebagai pengetahuan dini dalam proses pengambilan keputusan untuk tindakan preventif dalam hal mengantisipasi mahasiswa *drop out*, untuk meningkatkan prestasi mahasiswa, untuk meningkatkan kurikulum, meningkatkan proses kegiatan belajar dan mengajar dan banyak lagi keuntungan lain yang bisa diperoleh dari hasil penambangan data tersebut.

Berdasarkan uraian di atas, penulis terdorong untuk mengambil judul “ *Penerapan Data Mining dengan Menggunakan Metode Clustering K-Mean Untuk Mengukur Tingkat Ketepatan Kelulusan Mahasiswa Program Teknik*

Informatika S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang Tahun Ajaran 2012/2013”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan sebelumnya, maka penulis merumuskan masalah yang dihadapi adalah “Bagaimana menerapkan Data Mining dengan Menggunakan Metode Clustering K-Mean Untuk Mengukur Tingkat Ketepatan Kelulusan Mahasiswa Program Teknik Informatika S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang Tahun Ajaran 2012/2013.

1.3 Batasan Masalah

- Objek yang dikelompokkan adalah nilai Indeks Prestasi Mahasiswa.
- Data yang digunakan merupakan data IP mahasiswa dari semester 1 hingga IP semester 5
- Algoritma yang digunakan dalam melakukan clustering adalah algoritma K-Means

1.4 Tujuan Penelitian

Untuk melakukan pengelompokan masa studi para mahasiswa di program studi Teknik Informatika Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang Tahun Ajaran 2012/2013. yang nantinya akan

dapat dijadikan acuan bagi pihak universitas untuk melakukan atau membuat kebijakan bagi para mahasiswa khususnya yang berpotensi mengalami kemunduran masa studi dan berpotensi mengalami drop out

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah:

1. Bagi instansi pendidikan

Penelitian ini dilakukan agar instansi pendidikan dapat membantu mengambil keputusan yang dapat mengantisipasi dan memberi pemberitahuan mengenai mahasiswa yang potensial *drop out*.

2. Bagi penulis

Mengetahui bagaimana cara mengimplementasikan metode clusterig pada sistem komputer serta mengetahui tingkat keakuratan metode tersebut berdasarkan hasil perhitungannya pada penelitian ini.

3. Bagi perkembangan ilmu pengetahuan

Penelitian ini digunakan sebagai bahan referensi dalam menambah pengetahuan bagi peneliti lain yang memiliki relevansinya.

2. LANDASAN TEORI

2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. [

2.2 Clustering

Clustering adalah proses pengelompokkan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain. Perbedaan dan persamaannya biasanya berdasarkan nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak.

2.3 K-Means

K-Means merupakan algoritma clustering yang berulang. Algoritma K-Means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya cluster yang ingin dibentuk kemudian tetapkan nilai-nilai K secara acak, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau “means” hitung jarak setiap data yang ada

terhadap masing-masing centroid menggunakan rumus Euclidean hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid stabil

3. METODE PENELITIAN

1.1. Lokasi Penelitian

penulis mengadakan penelitian pada Fakultas Ilmu Komputer Program Studi Teknik Informatika Universitas Dian Nuswantoro Semarang.

1.2. Prosedur Pengumpulan Data

Prosedur yang digunakan peneliti antara lain :

Dalam penyusunan tugas akhir ini perlu adanya suatu metode tertentu yang akan digunakan dalam pengumpulan data yang diperoleh dengan cara sebagai berikut :

a. Studi Lapangan

Yaitu data yang secara langsung dapat diperoleh dengan meninjau dan mengamati secara langsung terhadap obyek yang diteliti melalui:

1. Wawancara (*Interview*)

Yaitu metode pengumpulan data yang

mengadakan tanya jawab langsung dengan atau lisan kepada pegawai tata usaha yang terkait dalam proses studi mahasiswa dan hasil studi mahasiswa.

2. Survey (Observasi)

Penulis mengamati atau melihat secara langsung pada fakultas ilmu komputer semarang yang menjadi objek penelitian, sehingga penulis mendapatkan gambaran secara lengkap.

b. Studi Pustaka

Yaitu penelitian dengan menggunakan dan mempelajari buku-buku maupun literatur-literatur yang berhubungan dengan masalah yang diteliti sebagai landasan teori bagi penulis .

1.3. Metode Analisis Data

Metode analisis data untuk teknik pembobotan dan perhitungan yang

dipergunakan dalam tugas akhir ini adalah metode *Clustering K Means*.

metode *Clustering K Means* digunakan karena metode ini dapat memecahkan masalah yang kompleks dimana kriteria yang diambil cukup banyak dalam memilih status mahasiswa. Derajat kepentingan pengguna dilakukan dengan pendekatan perbandingan berpasangan (*Pairwise Comparisons*).

Perbandingan berpasangan adalah membandingkan setiap elemen dengan elemen lainnya pada setiap tingkatan hirarki secara berpasangan, sehingga didapat nilai kepentingan elemen dalam bentuk pendapat kualitatif. Perbandingan berpasangan tersebut diulang untuk semua elemen dalam tiap tingkat. Elemen dalam bobot paling tinggi adalah pilihan keputusan yang layak dipertimbangkan untuk diambil.

4. IMPLEMENTASI DAN PENGUJIAN

4.1 Perhitungan Manual

Proses Perhitungan Teknik Clustering K-Means adalah sebagai berikut :

1. Menentukan Jumlah Cluster (K)
2. Menentukan Centroid tiap Cluster menggunakan

random (Pemilihan Acak)

untuk menentukan centroid awal

3. Lakukan proses Iterasi I dengan menggunakan perhitungan dari data dan hasil centroid awal
4. Lakukan proses Iterasi ke II menggunakan perhitungan data dan hasil centroid baru
5. Cocokkan hasil iterasi I dan II apabila hasilnya telah sama maka iterasi berhenti dan lakukan pengclusturan , apabila hasil masih beda lakukan iterasi berikutnya hingga menghasilkan hasil yang sama
6. Hasil perhitungan yang paling dekat (minimum) dengan Centroid merupakan anggota dari kluster tersebut

No	NIM	IP 1	IP 2	IP 3	IP 4	IP 5
1	A11.2012.06535	3,20	3,20	1,78	2,41	2,17
2	A11.2012.06537	3,30	3,26	3,35	3,08	1,50
3	A11.2012.06544	3,30	3,20	2,65	3,32	3,63
4	A11.2012.06545	3,20	2,90	3,05	3,13	2,63
5	A11.2012.06546	3,40	2,90	3,50	2,71	3,00
6	A11.2012.06547	3,60	3,91	3,83	3,79	3,88
7	A11.2012.06549	2,90	2,80	2,75	2,73	3,14
8	A11.2012.06550	3,00	3,20	2,96	2,62	2,71
9	A11.2012.06551	3,30	2,91	2,10	2,85	2,65
10	A11.2012.06552	3,00	3,20	3,09	3,21	3,25
11	A11.2012.06555	3,80	3,74	3,65	3,67	3,29
12	A11.2012.06556	4,00	3,83	3,83	3,92	3,75
13	A11.2012.06558	3,20	2,70	2,40	3,00	2,38
14	A11.2012.06559	3,20	2,33	2,35	2,70	2,35
15	A11.2012.06563	3,20	3,00	3,00	3,00	3,00
16	A11.2012.06564	3,30	3,30	3,26	3,17	2,75
17	A11.2012.06565	3,40	3,43	3,35	3,50	3,63
18	A11.2012.06566	3,40	3,30	3,17	3,29	3,50
19	A11.2012.06567	3,50	3,83	3,61	3,54	3,13
20	A11.2012.06568	3,00	3,10	3,00	3,00	3,13
513	A11.2012.07323	2,40	3,30	2,83	2,71	2,55

Data Mahasiswa

AWAL	IP 1	IP 2	IP 3	IP 4	IP 5
C0	4,00	3,83	3,83	3,92	3,75
C1	1,70	1,56	1,78	1,72	1,53
C2	3,10	3,09	2,57	3,27	2,88

Centroid

RUMUS ITERASI

Menghitung C0 =

$$\sqrt{(IP1 - C0IP1)^2 + (IP2 - C0IP2)^2 + (IP3 - C0IP3)^2 + (IP4 - C0IP4)^2 + (IP5 - C0IP5)^2}$$

Menghitung C1 =

$$\sqrt{(IP1 - C1IP1)^2 + (IP2 - C1IP2)^2 + (IP3 - C1IP3)^2 + (IP4 - C1IP4)^2 + (IP5 - C1IP5)^2}$$

Menghitung C2 =

$$\sqrt{(IP1 - C2IP1)^2 + (IP2 - C2IP2)^2 + (IP3 - C2IP3)^2 + (IP4 - C2IP4)^2 + (IP5 - C2IP5)^2}$$

Data IP Mahasiswa keseluruhan di hitung menggunakan rumus tersebut sehingga akan menghasilkan data untuk iterasi , setelah selesai perhitungan pada iterasi 0 , proses berikutnya adalah melakukan perhitungan iterasi 1 , namun sebelum melakukan iterasi 1 kita harus menentukan centroid baru , cara menentukan centroid baru ini berbeda dengan cara yang pertama yang prosesnya acak , untuk iterasi 1 cara menghitung centroid barunya adalah sebagai berikut :

C0 Baru : hasil rata – rata dari IP

Mahasiswa anggota C0 hasil Iterasi 0

C1 Baru : hasil rata – rata dari IP

Mahasiswa anggota C1 hasil Iterasi 0

C2 Baru : hasil rata – rata dari IP

Mahasiswa anggota C2 hasil Iterasi 0

Setelah mendapatkan nilai untuk iterasi 2 maka ulangi langkah yang sama dengan proses iterasi 1 , kemudian cocokkan hasil dari perhitungan iterasi 1 dan Iterasi 2 apabila telah menghasilkan hasil yang sama dan nilai centroid tidak berubah maka proses clusterisasi telah selesai

Namun apabila masih terjadi perbedaan hasil dan nilai centroid maka kita harus melakukan iterasi kembali dengan langkah yang pertama yaitu menentukan nilai centroid baru, kemudian lakukan perhitungan seperti pada proses yang pertama hingga menghasilkan perhitungan yang sama antara nilai centroid dan hasil perhitungan antar iterasi, jika telah menghasilkan nilai yang sama maka proses perhitungan iterasi dinyatakan telah selesai karena hasil perhitungannya telah sama dan tidak mengalami perubahan nilai antar iterasi

4.2 Hasil Implementasi

4.2.1 Nilai Centroid Menggunakan RapidMiner

Attribute	cluster_0	cluster_1	cluster_2
IP 1	3.440	2.742	3.008
IP 2	3.434	2.515	2.924
IP 3	3.366	2.387	2.945
IP 4	3.474	2.521	3.067
IP 5	3.358	2.369	2.876

4.2.2 Nilai Centroid Menggunakan Perhitungan Manual

BARU	IP 1	IP 2	IP 3	IP 4	IP 5
C0	3,44	3,43	3,37	3,47	3,36
C1	2,74	2,52	2,39	2,52	2,37
C2	3,01	2,92	2,95	3,07	2,88

4.2.3 Anggota Cluster Menggunakan RapidMiner dan Perhitungan Manual

Cluster Model

```
Cluster 0: 156 items
Cluster 1: 132 items
Cluster 2: 225 items
Total number of items: 513
```

Cluster 0	156
Cluster 1	132
Cluster 2	225
Total Data	513

Nilai centroid dan Anggota cluster telah memiliki nilai yang sama, maka proses iterasi telah selesai

4.2.4 Penentuan Anggota Clusteran

CENTROID	IP 1	IP 2	IP 3	IP 4	IP 5	rata - rata
C0	3.44	3.43	3.37	3.47	3.36	3.4
C1	2.74	2.52	2.39	2.52	2.37	2.4
C2	3.01	2.92	2.95	3.07	2.88	2.96

Berdasarkan nilai yang di tampilkan pada gambar diatas maka :

- C0 merupakan cluster yang anggotanya adalah data mahasiswa yang berpotensi menyelesaikan studinya dalam waktu yang lebih cepat
- C1 merupakan cluster yang anggotanya adalah data mahasiswa yang berpotensi mengalami kemunduran masa studinya dan berpotensi drop-out
- C2 merupakan cluster yang anggotanya adalah data mahasiswa yang menyelesaikan studinya dalam waktu normal 8 semester

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil analisis dan pengujian yang dilakukan pada bab sebelumnya, maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Pemahaman informasi mahasiswa yang potensial drop out penting untuk diketahui, pemahaman dapat dilakukan dengan mengungkapkan

pengetahuan yang dimiliki untuk memahami, memprediksi, dan pencegahan kegagalan adalah sangat penting bagi manajemen perguruan tinggi. Pengetahuan ini dapat digunakan dalam membantu pihak perguruan tinggi untuk lebih mengenal situasi para mahasiswanya, dan dapat dijadikan sebagai pengetahuan dini dalam proses pengambilan keputusan untuk tindakan preventif dalam hal mengantisipasi mahasiswa yang berpotensi drop-out,serta untuk meningkatkan prestasi mahasiswa, untuk meningkatkan kurikulum, meningkatkan proses kegiatan belajar dan mengajar yang bisa diperoleh dari hasil penambangan data

2. Dengan bantuan teknik *data mining* dan metode clustering, memungkinkan untuk menemukan informasi baru dari data yang berisi nilai IP dari mahasiswa dan pengolahannya dapat digunakan sebagai bahan acuan untuk memprediksi masa

studi mahasiswa

3. Hasil yang diperoleh diantaranya adalah kelompok mahasiswa yang berpotensi mengalami drop out, artinya mahasiswa-mahasiswa yang termasuk dalam anggota cluster yang berpotensi dropout dapat dijadikan acuan untuk mempermudah pengambilan keputusan terhadap mahasiswa yang bersangkutan.

5.2. Saran

Penulis menyarankan pengembangan penelitian lebih lanjut terhadap sistem Penerapan Data Mining dengan Menggunakan Metode Clustering K-Mean Untuk Mengukur Tingkat Ketepatan Kelulusan Mahasiswa Program Teknik Informatika S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang Tahun Ajaran 2012/2013 sebagai berikut:

1. Untuk mendapatkan hasil yang lebih variatif penelitian ini dapat juga di implementasikan pada seluruh program studi dan fakultas yang ada pada Universitas Dian Nuswantoro sehingga dapat membantu pihak universitas dalam mengontrol prestasi mahasiswa nya

2. Jika Penggunaan metode clustering ini kedepannya mampu membantu pihak universitas dalam mengetahui masa studi mahasiswa secara langsung , diharapkan metode ini dapat digunakan setiap tahun ajaran baru untuk mengevaluasi dan memanajemen masa studi dan prestasi mahasiswa Universitas Dian Nuswantoro sehingga akan tercapai kualitas yang baik dan terjaga

3. Untuk memperoleh akurasi sistem yang lebih tinggi dalam mengelompokkan data mahasiswa dapat di bandingkan dengan metode atau teknik data mining yang lain